Statistical treatment of the measurement data

Approved by the CEPI-CTS Working Group in St. Gallen on April 14, 2000
Revised by the CEPI-CTS Working Group in Grenoble on April 20, 2001
Revised and approved by the CEPI-CTS Working Group in Apeldoorn on April 23, 2004
Rev.6: updated 2015-10-06
Rev.7: updated 2020-01-01
Rev.8: updated 2020-10-10

0 General

The CEPI Comparative Testing Service (CEPI-CTS) is a proficiency testing scheme directed to the Pulp and Paper Industry and to its supplier and user industries. The CEPI-CTS covers more than 100 paper properties.

The CEPI-CTS operates by means of interlaboratory comparisons. According to the definitions given in ISO 13528:2015 ([1]) and ISO/IEC 17043:2010 ([2]), CEPI-CTS is a quantitative, simultaneous and continuous PT.

The performance of the participating laboratories is assessed against assigned values (CEPI Assigned Value) and their associated acceptable ranges of results (Warning and Action Limits) as determined by a preliminary interlaboratory comparison (pre-test round) of qualified laboratories (QLs): we are in the scenario of consensus values from expert participants).

This document describes in detail the statistical procedures to be used by the Co-ordinating Laboratory (CLs) in the analysis of the pre-test data, in the calculation of the CEPI Assigned Value and the associated Warning and Action Limits, and in the analysis and presentation of the client data. This analysis and reporting is automated within the CEPI-CTS spreadsheet and macro.

The basic principle is that the data provided by the QLs in the pre-test round is used to obtain estimates of the grand mean (which becomes the CEPI Assigned Value) and of the associated total standard deviation (which becomes the standard deviation for proficiency assessment) when sets of a given number of proficiency test items from a given batch are tested in different laboratories in the pre-test round. This total standard deviation s_{pt} is then used for estimating acceptable limits within which the majority of all laboratories should lie when testing test items from the same batch. These limits are presented as Warning Limits and Action Limits.

NOTE – The architecture and, partially, the methodology conform to those given in ISO 5725-1:1994 ([3]) and ISO 5725-2:2019 ([4]), and the same statistical test (Cochran and Grubbs) are used to identify outliers in the pre-test round. Nevertheless, the CEPI-CTS is *NOT* intended to calculate precision quantities or the uncertainty of a measurement method, but aims at assessing the performance of a participating laboratory.

1 Pre-test round

The CEPI Assigned Value and the CEPI Warning and Action Limits for each property covere by the CEPI-CTS are established by a pre-test round, in which a limited number of qualified laboratories (QLs) participate. In order to ensure the highest quality for these values, the statistical analysis of the pre-test data includes statistical tests to eliminate data which are considered to be outliers and which may not properly belong to the data population; however these outlier tests are only performed when results have been returned by more than ten qualified laboratories. The results of such pre-test round are presented in the Report A and Report A-Summary.

Simple formulae are used in the statistical analysis, based on the assumption that each laboratory makes the same number of measurements. Experience has shown that it is unnecessary to use the more complicated general formulae which permit different laboratories to use different numbers of test pieces.

1.1 Calculation of the mean and standard deviation for each laboratory

The following analysis shall be followed for each property and for each level without any attempt to combine data for different levels.

1.1.1 Calculate the mean for each laboratory as:

$$\bar{y}_i = \frac{1}{n} \sum_{k=1}^{n} y_{ik} \qquad [1]$$

where n is the number of test pieces in each set.

1.1.2 Calculate the standard deviation for each laboratory as:

$$s_i = \sqrt{\frac{1}{n-1} \sum_{k=1}^{n} (y_{ik} - \bar{y}_i)^2} \qquad [2]$$

1.2 Test for outliers with regard to within-laboratory standard deviation

In order to eliminate any laboratory which has an unreasonably high standard deviation, apply Cochran's test according to ISO 5725-2, with a rejection level of 1%. The test shall be applied only to the laboratory having the highest deviation. If the acceptance criterion for all remaining labs is still not met, the test should be applied once more (maximum twice).

The Cochran's test is not applied if the number of qualified laboratories is less than ten.

Cochran's statistic, *C*, is calculated as:

$$C = \frac{s_{i.\max}^2}{\left[\sum_{i=1}^{p} s_i^2\right]} \qquad [3]$$

where *p* is the number of participating laboratories.

Critical values for Cochran's test are given in *Table 1*. A laboratory is excluded if the value obtained is higher than the value given in the table.

Any laboratory excluded on the basis of too high a standard deviation shall be completely excluded from the subsequent analysis.

Table 1. Critical values for Cochran's test

	number of replicate measures *n*		
number of labs *p*	**n = 5**	**n = 10**	**n = 20**
2	0.9586	0.8674	0.7744
3	0.8335	0.6912	0.5841
4	0.7212	0.5702	0.4682
5	0.6329	0.4854	0.3910
6	0.5636	0.4229	0.3362
7	0.5080	0.3751	0.2952
8	0.4627	0.3373	0.2636
9	0.4251	0.3067	0.2382
10	0.3934	0.2813	0.2173
11	0.3670	0.2606	0.2000
12	0.3428	0.2419	0.1852
13	0.3236	0.2271	0.1726
14	0.3055	0.2134	0.1616
15	0.2882	0.2002	0.1519
16	0.2748	0.1904	0.1435
17	0.2616	0.1807	0.1359
18	0.2497	0.1719	0.1291
19	0.2388	0.1639	0.1229
20	0.2288	0.1567	0.1173
21	0.2200	0.1503	0.1123
22	0.2119	0.1444	0.1077
23	0.2043	0.1390	0.1034
24	0.1970	0.1338	0.0995
25	0.1907	0.1292	0.0959
26	0.1846	0.1249	0.0926
27	0.1788	0.1208	0.0894
28	0.1734	0.1170	0.0865
29	0.1683	0.1134	0.0838
30	0.1635	0.1100	0.0812

1.3 Calculation of the preliminary grand mean

Calculate the grand mean as:

$$\bar{\bar{y}} = \frac{\left(\sum_{i=1}^{p} \bar{y}_i\right)}{p} \qquad [4]$$

1.4 Calculation of the preliminary total standard deviation

Calculate the total standard deviation of the grand mean as:

$$s = \sqrt{\frac{1}{p-1} \sum_{i=1}^{p} \left(\bar{y}_i - \bar{\bar{y}}\right)^2} \qquad [5]$$

Note that this is not the analytical standard deviation between laboratories. It includes not only the between-laboratory deviation but also the deviation within laboratories and the variation between the test pieces.

1.5 Test for outliers with regard to laboratory mean value

In order to eliminate any laboratory which presents a mean value which is unreasonably far from the grand mean, apply Grubbs's test according to ISO 5725-2 to check whether a high or low mean value (the one which is farthermost from the grand mean) is to be regarded as an outlier, with a rejection level of 1%. If the acceptance criterion for all remaining labs is still not met, the test should be applied once more (maximum twice).

The Grubbs's test is not applied if the number of qualified laboratories is less than ten.

Grubbs's statistic, *G*, is:

$$G = \frac{\left(\bar{y}_i - \bar{\bar{y}}\right)}{s} \qquad [6]$$

Critical values for Grubbs's test are given in *Table 2*. A laboratory is excluded if the value obtained is higher than the value given in the table.

Any laboratory excluded on the basis of a outlying mean shall be completely excluded from the subsequent analysis.

Table 2. Critical values for Grubb's test

number of labs *p*	G	number of labs *p*	G
3	1,155	15	2,806
4	1,496	16	2,852
5	1,764	17	2,894
6	1,973	18	2,932
7	2,139	19	2,968

number of labs *p*	G	number of labs *p*	G
8	2,274	20	3,001
9	2,387	21	3,031
10	2,482	22	3,060
11	2,564	23	3,087
12	2,636	24	3,112
13	2,699	25	3,135
14	2,755	26	3,157

1.6 Calculation of the CEPI Assigned Value (grand mean), x_{pt}

After elimination of any outliers, calculate the CEPI Assigned Value, i.e the grand mean of the remaining laboratories, as:

$$x_{pt} = \frac{\left(\sum_{i=1}^{p'} \bar{y}_i\right)}{p'} \qquad [7]$$

where p' is the number of laboratories accepted on the basis of the Grubbs's test. Such grand mean is the CEPI Assigned Value.

1.7 Calculation of the mean standard deviation within laboratories, s_w

After elimination of any outliers calculate the mean within-laboratory standard deviation (CEPI-SD Within) of the remaining laboratories as:

$$s_w = \sqrt{\frac{\left(\sum_{i=1}^{p'} s_i^2\right)}{p'}} \qquad [8]$$

Note that this is not the repeatability standard deviation within laboratories but is primarily a measure of the variation between the test pieces.

1.8 Calculation of the standard deviation for proficiency assessment (total standard deviation), σ_{pt}

After elimination of any outliers, calculate the standard deviation for proficiency assessment for the remaining laboratories (CEPI Total SD) as:

$$s_{pt} = \sqrt{\frac{1}{p-1}\sum_{i=1}^{p'}\left(\bar{y}_i - x_{pt}\right)^2} \qquad [9]$$

This total standard deviation used for proficiency assessment includes all sources of variation such as the between-laboratory variation, and the variation between the proficiency test items. Variations due to transport instability and instability over time are

considered negligible, thanks to the precautions taken in the preparation and shipment of test items.

1.9 Calculation of the acceptable range of results (CEPI Warning and Action Limits)

The warning and action limits are calculated using Student's t distribution value.

Calculate thus the warning limits as:

$$WL = x_{pt} \pm w \times s_{pt} \quad [10]$$

where w = 2.0 is the Student's t value for a probability of .05 and p=999

And the action limits as:

$$AL = x_{pt} \pm a \times s_{pt} \quad [11]$$

where a = 2.6 is the Student's t value for a probability of .01 and p=999

NOTE – The comparison of the result of a participant with the limits as defined above is equivalent to the use of the *z* score. If the *z* score is used, ISO/IEC 17043 and ISO 13528 suggest action limits be set as the assigned value ± 3 times the standard deviation for the proficiency assessment (see ISO 13528 §9.4.2). However, years of experience has lead CEPI-CTS to use the narrower action limits because the higher natural variance observed in testing paper-based materials can often cause high variances which in turn lead to excessively wide limits.

2 Client's round

The Report H presents a statistical elaboration of the data for all laboratories participating in the client's round. It can include the results of the QLs if the number of clients is less than ten. If this is the case, this should be mentioned in the Report H.

The Report H shows the grand mean and standard deviation of the client population together with the number of outliers (clients reporting results outside the CEPI Action Limits).

The results in the Report H are also presented in the form of a histogram showing the distribution of all client results, i.e. including those outside the action limits.

2.1 Calculation of the mean and standard deviation for each laboratory

Calculate the mean and standard deviation for each participating client laboratory according to equations [1] and [2].

Record the number of participating clients *q*.

Determine the number of clients for which the mean is outside the CEPI Action Limits (outliers).

2.2 Calculation of the client grand mean, X_C

Calculate the grand mean for all the clients including all outliers (outside the CEPI Action Limits) as:

$$X_C = \frac{\left(\sum_{i=1}^{q} \bar{y}_i\right)}{q} \qquad [12]$$

where q is the total number of clients.

2.3 Calculation of the mean standard deviation within laboratories

Calculate the mean within-laboratory standard deviation as:

$$s_{Cw} = \sqrt{\frac{\left(\sum_{i=1}^{q} s_i^2\right)}{q}} \qquad [13]$$

Note that this is not the repeatability standard deviation within laboratories but is primarily a measure of the variation between test pieces.

2.4 Calculation of the total standard deviation

Calculate the total standard deviation of the client grand mean as:

$$s_C = \sqrt{\frac{1}{q-1}\sum_{i=1}^{q}\left(\bar{y}_i - X_C\right)^2} \qquad [14]$$

This total standard deviation includes all sources of variation such as the between-laboratory variation, and the variation between the proficiency test items. Variations due to transport instability and instability over time are considered negligible, thanks to the precautions taken in the preparation and shipment of test items.

2.5 Construction of the Report H histogram

For Report H, draw a histogram showing the distribution of the mean values of all the client laboratories.

The number of classes in the histogram shall be 13 if the number of clients exceeds 50 and 9 if the number of clients is less than or equal to 50.
If the number of clients is less than 10, data from QL's pre-test round will be added to the histogram, in order to obtain a meaningful distribution. If this is the case, this should be mentioned in the Report H.

3 Literature

([1]) ISO 13528:2015 Statistical methods for use in proficiency testing by interlaboratory comparison.

([2]) ISO/IEC 17043:2010 Conformity assessment - General requirements for proficiency testing

([3]) ISO 5275-1:1994 Accuracy (trueness and precision) of measurement methods and results – Part 1: General principles and definitions

([4]) ISO 5275-2:2019 Accuracy (trueness and precision) of measurement methods and results – Part 2: Basic method for the determination of repeatability and reproducibility of a standard measurement method.